

This work is licensed under a
Creative Commons Attribution-NonCommercial-
NoDerivs 3.0 Licence.

To view a copy of the licence please see:
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

(832)

IDS LIBRARY
RESERVE COLLECTION

INSTITUTE FOR DEVELOPMENT STUDIES

UNIVERSITY COLLEGE, NAIROBI

Discussion Paper No. 48

HOUSEHOLD EXPENDITURE IN NAIROBI: A STATISTICAL
ANALYSIS OF CONSUMER BEHAVIOR

Benton F. Massell and

Judith U. Heyer

April 1967

Discussion Papers are working papers that report on tentative or interim research results. They are distributed for discussion purposes only.

Household Expenditure in Nairobi: A Statistical
Analysis of Consumer Behavior

Benton F. Massell and Judith U. Hayer

Introduction

This study presents estimates of relationships between household expenditure patterns and other major variables, for use in predicting urban household demands for selected groups of goods and services. The estimates are obtained from a sample of 300 middle-income African households in Nairobi.

Household budget surveys have been undertaken by many Governments in Africa for the compilation of cost of living indices. They are relatively cheap and easy surveys to do, and there are now several available within East Africa for different consumer groups. Although designed primarily for cost of living purposes, these surveys also represent a useful source of information on consumer demands, the only source available in most African countries at present. It is stated that the survey used in this study was planned in order to estimate consumer demand rather than cost of living indices,¹ but, like the others, it does not appear to have been organised for this purpose at the data collection stage. Much of the detail required for cost of living indices is the same as that required for the estimation of final demands, but there are one or two important differences.

Final demands can be predicted if the relationship between expenditure patterns and some associated variable subject to clearly defined future trends is known. The most obvious variable closely related to expenditure patterns, whose trend can be predicted, is income; the most useful prediction information that can usually be obtained from household budget surveys is on income elasticities. One normally considers other possible variables as well: age, education, or household size, for example. But income tends to be the variable most closely related to expenditure patterns, and income elasticities are most satisfactory anyway for

1. Kenya Government, The Pattern of Income, Expenditure and Consumption of African Middle Income Workers in Nairobi, July 1963, 1964, p. 1.

prediction purposes, income being subject to easily estimated future trends.

Unfortunately, household budget surveys, particularly those designed to estimate weights for cost of living indices, do not usually have reliable income figures. It is therefore often necessary to estimate expenditure rather than income elasticities, as we have had to do here. One can make assumptions about the future trend of total expenditure as opposed to income, and predict consumer demands on the basis of this. But it would be preferable to have more reliable income information, which would enable more accurate demand predictions and some estimation of savings trends as well.

For economic planning, we want to know the future pattern of final demands in the economy, and to predict these we need a series of household budget surveys covering representative consumer groups for the country as a whole. In Kenya, so far, we only have surveys for a few consumer groups, and we are not yet in a position to aggregate over the whole economy.

We have estimated expenditure elasticities for major items of consumer expenditure for a particular group of wage earners in Nairobi. We cannot claim that this covers Nairobi wage earners as a whole, or other urban groups. However, projections need to be made, and our elasticities will provide some (however incomplete) basis for such projections until further information is available; we have accordingly attempted a very rough exercise that assumes that our elasticities do hold for Kenya's urban population as a whole. This shows the use that can be made of these figures, particularly in future when the data coverage is improved.

The Variables

We are principally interested in the relationship between a household's expenditure pattern and its income. However, there are problems in the data we are using with the definition of income: for example, in the treatment of fringe benefits, overtime pay and imputed rent. Moreover, although income should refer to the total income of all household members, there is reason to believe that in many households only the

respondent's income was recorded.¹ Income as measured in the survey does not correspond to the economic concept of income, which makes it hazardous to use it as an explanatory variable. What is commonly done in such studies, and what we have chosen to do here, is to use total household expenditure as an explanatory variable in place of income. Thus our principal concern centers on the relationship between a household's expenditure pattern and its total expenditure.

By expenditure pattern, we mean the allocation of total expenditure among 26 expenditure groups, discussed in more detail below. These expenditure groups form a set of dependent variables which are related to total expenditure and other explanatory variables. By including other explanatory variables in the regressions, the effect of total expenditure is estimated net of the effects of these other variables. It is also of some interest to estimate the effects of these other variables themselves on expenditure patterns.

Five explanatory variables, in addition to total expenditure, have been considered: (1) household size; (2) payday; (3) acreage in the rural areas; (4) education of the respondent; and (5) in the case of expenditure on housing, whether the housing is provided by the employer or not. These variables are discussed below.

Our analysis is based on the Kenya Government household budget survey of African middle-income workers in Nairobi in July 1963. A summary of the data was published in 1964.² The number of households included in the survey was 324. A comparison between the sample and the total population, given by the 1963 labour enumeration, suggests that the sample is biased towards the upper income groups in the range covered, and toward Government as opposed to private sector employees. Forty seven percent of Nairobi's employees in this income range, but 58 percent in the sample, work for Government. The different income groups have sampling fractions rising progressively from 1.7 percent in the lowest group to 5.5

1. It is likely that the difference between total household income and recorded income would be greatest for households with a low recorded income, introducing a systematic bias into the results.

2. Kenya Government, op. cit.

percent in the highest, showing a substantial bias towards the higher income groups. It is important to bear this in mind in interpreting results.

We rejected households for which recorded observations appeared inconsistent or extreme, because we felt that failure to do so would distort the results.¹ We eliminated 24 households in this way, leaving 300 as a convenient number for grouping purposes. The households rejected were fairly evenly spread through the income range, although we did reject a higher number from the lowest income group.

Income does not appear directly in the analysis, but it is used to group the observations as explained later. Its definition is of some importance. The survey definition of income included all earnings (in cash or in kind), allowances and fringe benefits from regular employment, but not the value of any housing subsidy. It also included the earnings of other household members, which appeared to be recorded very irregularly. The range of income covered by the survey was from Shs.335/- to Shs.1399/- per month, with an average of shs.641/-. We defined income somewhat differently. We excluded the income of members of the household other than the head, because of the apparent recording irregularity. There were very few households for which additional income was recorded, and the survey forms for this item looked inadequate. We included the value of any housing subsidies given by employers. Forty five percent of the households in the sample had housing provided by their employers, so this made a considerable difference to the income figures. Our adjusted income ranged from Shs.335/- to Shs.1305/- with an average of Shs.655/-. Average total expenditure was 87.3 percent of income, in our figures which gives some idea of the order of magnitude of the savings element, although we would not attach too much weight to this for reasons already noted.

1. We excluded a household in which a car was purchased during the survey month, for example. This expenditure of Shs.8000/- was greater by a factor of 20 than the next highest expenditure on transportation.

Expenditure was defined in the survey to include items paid for in cash, but not items bought on credit. This is an unsatisfactory definition for the purposes of economic analysis, and may bias the estimated elasticities. A normal expenditure item may not appear because it is bought on credit, and recorded expenditure patterns may be distorted for this reason. However, there was no way in which we could alter this shortcoming in the data.

Average expenditure on the different items is shown in Table 1, where our figures are given alongside the published survey figures for comparison. We defined total expenditure to include all major items listed there. We regarded remittances as an item of expenditure, as they are usually maintenance payments for dependents living in the rural areas, and only a fraction is likely to go into investment rather than direct consumption. School fees were omitted because they were not strictly comparable. All expenditure figures except school fees relate to actual expenditure during the recording month, July 1963. School fees, as we defined them, are one twelfth of annual expenditure, which seems to us more meaningful than actual expenditure in July. School fees are usually paid on a termly or annual basis, not each month. The average July figure, given in the survey, is Shs.7/25; ours is Shs.16/50.

A few small expenditure items were also excluded from our total expenditure figures. These were rates, water, building materials (for the rural areas), and a miscellaneous item including gifts, licences, union dues, and legal fees. These together were responsible for an average of less than 5 percent of total expenditure. We have also omitted expenditure on fringe benefits and taxes.

The major food expenditure items were grouped into 14 categories, plus total food. Our grouping occasionally combines items that are separate in the Government figures in order to eliminate zero observations, which create difficulties in estimating the elasticities. The non-food items include 2 non-durables, 2 durables, 6 services and remittances. Among these there are a few that need further definition. Household equipment includes furniture and utensils; household operation includes laundry, cleaning,

shoe repairs, and servants' wages; health includes both medical expenses and personal toilet items; recreation includes such things as entertainment, books, stationery, magazines, and records. Our transport figure is considerably lower than that of the Government because of the excluded observation mentioned in the footnote earlier.

Housing was treated in a very unsatisfactory way in the Government survey. Housing expenditure given in the Government report relates to the actual rent paid in cash, excluding the value of any housing subsidy. Our figures are economic rents: net rent paid in cash plus the value of any housing subsidy contributed by the employer.

Apart from the exceptions noted, our expenditure averages are very close to those given in the Government report, as can be seen in Table 1.

The numbers of adults and children under 16 years in the household were recorded separately. We combined the two to form a measure of consumption units, in which children were given half the weight of adults. The use of consumption units is never entirely satisfactory; weighting should probably differ for different items. But other household budget studies provide no conclusive information as to the best weighting system to use. The average number of adults per household in our sample was 2.5, children 2.2, and consumption units 3.6. Total earnings and total expenditure per consumption unit were Shs.181/- and Shs.158/-. The consumer unit index is used as a measure of household size in the regressions.

Thirty nine percent of the respondents had completed 9 years or more of education, and 48 percent were continuing their education at the time of the survey. This is a remarkably high figure for further education among adults. It seems likely that the education of the respondent will influence his expenditure pattern. In acquiring education, a person's tastes change. Education also affects a person's income prospects. In part this may be reflected in a higher level of income as recorded in the survey, but it may also reflect an expectation of greater future income prospects. Of two persons with the same present income, the more educated person may expect a higher future income, and may accordingly spend differently. In the analysis we used a dummy variable for education,

distinguishing persons with 9 or more years of formal schooling from those with less than 9 years.

Marital status was recorded, showing that 84 percent of the respondents were married. But it is difficult to attach a great deal of significance to this. Many of the married men had their wives and families living in the rural areas, and with some the connection was probably remote. Many Nairobi wage-earners have informal liaisons in the town, too, and there may be little difference between this and legal marriage. Preliminary investigations suggested that this variable did not significantly affect expenditure patterns.

The respondent's tribe was also recorded. In the sample, 31 percent were Kikuyu, 23 percent Luo, 19 percent Baluhya, 17 percent Kamba, and 10 percent other tribes. Here again an examination of the figures suggested that tribal differences in expenditure patterns were not sufficiently marked to be worth further analysis.

Acreages of land owned in the rural areas were given, and 41 percent of the respondents said they had 5 acres or more at home. It is difficult to know how accurate these figures are, as there is no check on what the respondent says. Although there is no strong reason to expect acreage to influence the respondent's expenditure pattern, acreage owned might serve either as a wealth variable or as an index of the strength of ties to the home area. Accordingly, we included a dummy variable for acreage, distinguishing those with at least 5 acres of land from those with less than 5 acres, including those with none.

There was some information on dependents not living with the respondent, but we would have needed more detail to treat this satisfactorily as an explanatory variable. We did not attempt to use it in the analysis.

Occupations were grouped in the survey, and we found it difficult to attach much meaning to the official classification. The breakdown consisted of: clerks, 39 percent; artisans, 17 percent; typists, cashiers, operators, telephonists, 11 percent; policemen, firemen, hospital workers 6 percent; drivers, 3 percent; supervisors, headmen, 3 percent, others 10 percent; and not stated, 11 percent. Occupation could have been an interesting explanatory variable if more suitably classified.

The average length of service, $7\frac{1}{2}$ years, was surprisingly high, particularly in view of the comparatively recent Emergency period, during which most of Nairobi's Kikuyu were repatriated to the rural areas. The average of $7\frac{1}{2}$ years includes many extremely long records of service in the same job. We investigated length of service as a possible explanatory variable but preliminary results suggested that it was not important.

The survey was conducted in July. Some respondents employed by the Government were paid earlier than usual because June is the end of the Government financial year. Thus a man who is normally paid at the end of the month may have been paid as much as 10 days earlier than usual in June. If an individual's expenditure pattern is influenced not only by his income, but also by cash receipts, then his expenditure pattern will differ according to when he was paid. We have therefore introduced a dummy variable for payday, taking on the value of unity if the respondent was paid prior to June 28th and zero otherwise. Thirty eight percent of the sample were paid early according to this definition, 10 percent by as much as 10 days. One would expect those paid earlier to have a different pattern of expenditure during the survey month.

Finally, in examining expenditure on housing, we added a dummy variable to distinguish between housing provided by the employer and housing provided directly by the respondent himself. We felt that this would make a difference to the household's housing expenditure.

Functional Forms

Two alternative functional forms were used: the double log and the ratio semilog. The double log function has been used in many expenditure studies and needs little explanation. It is written

$$\log(E_i) = a_{0i} + a_{1i}\log(E) + u_i \quad (1)$$

where E_i = expenditure on item i , E = total expenditure, the a 's are parameters to be estimated, and u_i is a disturbance term. Additional explanatory variables can be added to equation (1).

The principal advantage of (1) is its simplicity. The expenditure elasticity is simply the coefficient a_{1i} . The function has been shown to give a good fit in a wide variety of situations and it makes reasonable economic sense. The chief disadvantages are that it assumes constant elasticities and it fails to satisfy the additivity criterion: i.e. the weighted average of the estimated elasticities does not necessarily equal unity.

A function that is somewhat more difficult to work with, but which has more desirable properties, is the ratio semilog, written

$$\frac{E_i}{E} = b_{0i} + b_{1i}\log(E) + v_i \quad (2)$$

Equation (2) does satisfy the additivity criterion. Only if this criterion is satisfied does the level of disaggregation have no effect on the estimates.

The elasticity for (2) can be written

$$e_i = 1 + \frac{b_{1i}}{\frac{E_i}{E}} \quad (3)$$

Equation (3) can also be written

$$e_i = 1 + \frac{b_{1i}}{b_{0i} + b_{1i}\log(E)} \quad (3a)$$

where the relationship between e_i and E is more explicit. It can be seen that if $e_i = 1$, then e_i is not a function of E . However, if $e_i > 1$, then as E increases, e_i declines, tending toward 1. This is the case of a luxury good, which will form an increasing proportion of total expenditure as total expenditure increases. For $e_i < 1$ ($b_{1i} < 0$), as E increases e_i declines toward zero. There is a saturation level at which e_i -- and the marginal propensity to consume the item -- become zero; this level is at the point

$$\log(E) = - \left[1 + \frac{b_{oi}}{b_{li}} \right] \quad (4)$$

For increases in E beyond this level, e_i becomes negative and continues to decline. In general, the rate of decrease in e_i with respect to an increase in E is greater the more by which e_i differs from unity.

Deflation by Household Size

In many studies expenditure and income are expressed on a per capita or per consumer unit basis. We have decided against this and have instead included household size as a separate explanatory variable. This permits the possibility of economies or diseconomies of scale in expenditure. If the effect of household size depends on the level of household income, then deflating by household size will yield misleading results.

A second reason for treating household size as a separate variable is that deflation may lead to spurious correlation. It can be shown that this will be the case if a normally distributed regressand and normally distributed regressor are each deflated by an independently normally distributed variable.

Estimation

The use of ordinary least squares to estimate the coefficients in either (1) or (2) will yield biased estimates of the expenditure elasticities: A recent study has shown that this bias may be considerable, as much as 50% for some expenditure items.¹ However, the expenditure elasticities can be estimated consistently if the observations are grouped by income, income thus serving as an instrumental variable. The results then hold whether income is measured with error or not. All that is required is that income be statistically independent of the disturbance term in equation (1) or (2), a condition that can be assumed to hold here.

1. See N. Liviatan, Errors in Variables and Engel Curve Analysis, Econometrica, 1961.

Besides affording consistent estimates, working with grouped data had two further advantages. First, in view of the limited computer facilities available in Nairobi, grouping saved considerable time and expense. Our sample of 300 households was reduced, for computational purposes, to 60 groups of 5 households each. Second, the data contain a number of zero values which cause estimation problems. The zeros can be eliminated by grouping. The logarithm of zero is not defined, so one cannot use the double log function when there are zero observations. Even with the ratio semilog function there is a problem: the elasticity is usually calculated at the geometric mean of the variables, and this is zero if any of the observations is zero.

There are several ways to deal with zero observations. One can substitute a small positive number for each zero value. This may be acceptable if the number of zeros is sufficiently small, but is unsatisfactory otherwise. The estimated elasticity depends on what constant is chosen, and the choice is arbitrary. A second alternative is to use a functional form that is arithmetic in the dependent variable. But such functional forms have other less desirable properties than the two we have used. A third alternative is to group the data, as we have done, with enough households in each group for none of the group means to be zero.¹

Regression Results: Double Log

We used the double log function to estimate elasticities first. We then took the ratio semilog function and compared the elasticity estimates this gave with the double log estimates.

The first set of regressions includes total expenditure and household size as the explanatory variables. The estimated coefficients appear in Table 2. The last column of this table shows the determination coefficients, which measure the proportion of the variance in the log of the dependent variable that is explained by the regression as a whole.

1. The only expenditure item which had a zero value after grouping was eggs. This zero was altered to unity.

The determination coefficients range from .006 (school fees) to .658 (total food). A single asterisk against the determination coefficient indicates that the regression as a whole is significant at the .05 level, a double asterisk the .01 level, using an analysis of variance test. Of the 26 regressions, 21 are significant at the .01 level, and another 2 at the .05 level. Only 3 are not significant at all: eggs, meals away, and school fees.

The total expenditure elasticities appear in column 1 of the table. Using a two-tail t-test, total expenditure is statistically significant at the .01 level in 14 regressions, and at the .05 level in an additional 4 regressions. The items with higher expenditure elasticities naturally tend to be more significant, with the exception of eggs.

The elasticity of expenditure on total food is relatively low for a developing country, .483. However, urban middle-income households are likely to have a lower expenditure elasticity for food than rural or lower income households in the country as a whole.

The results suggest that maize is the only inferior good with a negative expenditure elasticity, although a 5 percent confidence interval includes small positive values for this elasticity as well. The very low expenditure elasticity for pulses, together with its high standard error, suggests that this elasticity could well be negative too. All other items have estimated elasticities between zero and unity. Cereals (.227), sugar (.285), and tea/coffee/soft drinks (.287) all have quite low elasticities, cereals and tea/coffee/soft drinks not significantly different from zero; meals away (.818), rice (.789) and fats (.599) have expenditure elasticities that are relatively high.

As would be expected, nonfood items tend to have substantially higher expenditure elasticities all round. Three items, school fees, fuel, and household operation, have estimated elasticities substantially below unity. School fees are typically recognised as income inelastic, and it is not too surprising to find fuel and household operation in the 'essentials' class as well.

The estimated elasticities for the other nonfood items all exceed unity. They are all significantly greater than zero at the .01 level, and some (equipment, clothing) are significantly greater than 1. These are all thus luxury goods and services, expenditure on which increases more than proportionately with increases in total expenditure. The items with the highest expenditure elasticities are equipment (1.948), clothing (1.644), and remittances (1.419).

We now turn to the household size elasticities, also given in Table 2. In the 26 regressions, 11 of the household size elasticities are significant at the .01 level, and an additional 5 are significant at the .05 level. In some cases, (pulses and maize), a high household size elasticity is associated with a low expenditure elasticity. For these items an increase in household size operates as a reduction in per capita income, forcing the family to consume relatively more of the inexpensive foods. A negative household size elasticity is associated with meals away, transport, recreation, clothing, and housing, all items with high expenditure elasticities. Apparently an increase in household size places demands on the household that compete with the demand for these luxury and semi-luxury items, forcing the household to reduce its expenditure on them. The relatively low household size elasticity for total food (.357) suggests a high degree of substitution of cheaper for more expensive foodstuffs as household size increases, resulting in a reduction of food expenditure per consumer unit. Larger households, (allowing for differences in income), tend to spend relatively more on pulses and maize, less on rice, wheat and fats, and absolutely less on eggs and meals taken out.

The household size elasticity for remittances is highly negative. There are two possible explanations for this. (1) Household size may be highly negatively correlated with dependents not living with the respondent, and remittances may be closely related to the number of these dependents living in the home area. (2) A wage earner living in a larger household may have greater expenditure obligations and may accordingly have less money left over to send home. These two explanations are not mutually exclusive, of course.

Also shown in Table 2 is a regression of housing against total expenditure, household size, and whether employer-provided or not. When the dummy variable is included to distinguish between employer-provided and employee-provided housing, the results alter considerably. The employer-provided coefficient is significantly less than zero at the .01 level, indicating that employer-provided accommodation is substantially less expensive than the housing chosen by an individual on his own. This is interesting and suggests that the employer tends to provide accommodation of lower quality than that chosen by the individual when employer-provided housing is not available.

The expenditure elasticity is less when taken net of the employer-provided variable (.871 as opposed to 1.076). This suggests that those for whom housing is provided tend to have a lower than average income (and thus total expenditure). What appears to be a high expenditure elasticity is in part the result of a shift from employer-provided to employee-provided housing. It is a matter of judgement which of the two elasticities is appropriate for particular prediction purposes.

The second set of regression results is based on three explanatory variables: total expenditure, household size and payday. Payday, as noted above, is expressed as a dummy variable, distinguishing those respondents paid before the 28th of June from those paid on the 28th or later. The estimated coefficients appear in Table 3. None of the payday coefficients is significant at the .05 level: indeed, very few of the t-ratios exceed unity. Moreover, the expenditure and household size elasticities in Table 3 are very little different from those in Table 2. This suggests that payday is of virtually no importance in explaining inter-household differences in expenditure patterns.

The statistical insignificance of payday is curious, suggesting that the expenditure pattern of an individual is not altered by his being paid early. One explanation for this may be that employers extended credit to employees who were paid early. It may also be that grouping the households by income obscures the effect of payday which can be regarded as a measurement error in the total expenditure variable.

The third set of regressions includes total expenditure, household size and acreage as explanatory variables. Acreage is measured as a dummy variable distinguishing between respondents with 5 acres or more of land in the rural areas, and respondents with less than this. None of the acreage coefficients is significant at the .05 level. However, it is worth noting that those owning larger acreages of land tend to consume less of nearly all the food items (the exceptions being pulses and spices) and of food as a whole. Larger landowners tend to spend more on selected nonfood items: fuel, equipment, household operation, and health, and to send more home as remittances.

None of the expenditure or household size elasticities is greatly altered by the inclusion of acreage in the regressions. We do not therefore give further consideration to this variable.

The fourth set of regressions to be examined includes total expenditure, household size and the education of the respondent. As already noted, education is treated as a dummy variable distinguishing respondents with at least 9 years of formal education from those with less than this. The education coefficient is significantly different from zero at the .05 level in only one regression: rice. The high degree of intercorrelation between education and total expenditure is probably responsible for the large standard errors and the consequent insignificance of this variable.

The high standard errors of the education coefficients make it hazardous to do more than speculate on the effect that education has on expenditure patterns. However, we note that the better educated tend to spend more on wheat, slightly more on total food, less on household equipment and household operation, and more on school fees. They also send less home as remittances, perhaps because they have weaker ties with the rural areas, or alternatively, stronger financial commitments in the town.

The inclusion of education in the regressions does have a substantial effect on some of the estimated expenditure elasticities. When taken net of education, expenditure elasticities are lower for rice and vegetables, and higher for equipment, health and remittances. Some of the

estimated household size elasticities (notably rice and remittances) are also altered when estimated net of education.

Before turning to the ratio semilog regressions, we summarise briefly the results so far. Total expenditure and household size are statistically significant in explaining interhousehold variation in most of the individual expenditure variables. Payday and acreage appear to add nothing to the explanation of expenditure patterns. Education is significant in only 1 of the set of 26 regressions. Its insignificance is probably due, at least in part, to multi-collinearity. The inclusion of education does change some of the expenditure and household size elasticities, substantially, though, unlike the inclusion of acreage or payday. The distinction between employer-provided and employee-provided housing is significant in the housing regression. Using the double log function, the first set of coefficients appear the most reliable for predicting demand.

Regression Results: Ratio Semilog

Because of the additional time and expense involved in running regressions with the ratio semilog function, we have calculated only one set of estimates: those based on the original pair of explanatory variables, total expenditure and household size. These results appear in Table 6.

The determination coefficients in the ratio semilog regressions need to be adjusted to make them comparable to those in the double log regressions. This is done by calculating, for each semilog regression, the proportion of the variance in $\log(E_i)$ explained by the regression. This is more meaningful than comparing the proportion of the variance in E_i/E explained by the semilog regression with the proportion of the variance in $\log(E_i)$ explained by the double log one. The proportion of the variance in E_i/E explained by the semilog regression can be zero even if E_i is highly correlated with E . The R^2 's given in Table 6 are for $\log(E_i)$ in the semilog regression, and are thus directly comparable with those in Table 2 for the first set of double log regression results.

(This calculation of determination coefficients for the ratio semilog function is a time consuming task, and has not yet been completed. These figures will appear in the final version of this study, to be completed within the next few weeks. For the purposes of the present draft, we merely note that preliminary evidence suggests that the determination coefficients for the double log and semilog regressions are very similar.)

The elasticities and their standard errors appearing in Table 6 have been calculated at the geometric mean of the variables, from equation (3) above. Twelve of the expenditure elasticities are significant at the .01 level, and an additional 5 at the .05 level. This compares with 14 and 4, respectively, in Table 2 with the double log function. For household size, 10 coefficients are significant at the .01 and one at the .05 level, compared with 11 and 5 in Table 2. Thus the two explanatory variables are individually slightly less significant in the ratio semilog than in the double log regressions.

The elasticities in Table 6 are very similar to those in Table 2. Two expenditure elasticities (pulses and school fees) and one household size elasticity (equipment) have switched from plus to minus; and one household size elasticity (eggs) has switched from minus to plus; but none of these elasticities is significantly different from zero.

The results suggest that the choice between the two alternative functional forms has little affect on the estimates; this is especially so with respect to those estimates that are statistically significant. The double log function appears to give a marginally better fit to the data. Moreover, the double log function is easier to work with, so for tentative demand projections we shall use the figures from the first set of regressions.

Economies of Scale

If we look at the sum of the household size and expenditure elasticities in the double log function we can see how important economies or diseconomies of scale in consumption are. If, for some expenditure item, the sum of the elasticities is less than one, then per capita expenditure on the item falls as household size rises: there are

economies of scale in the consumption of the item. Similarly, if the sum of the elasticities exceeds one, there are diseconomies of scale. One of the major objections to deflating by household size is removed if there are no economies or diseconomies of scale.

We tested for economies of scale in each of the 26 expenditure groups. Using a two-tail t-test, we tested the null hypothesis that the sum of the elasticities equals one. In only two cases was the test significant at the .05 level. The sum was significantly less than one for school fees, indicating economies of scale, and significantly greater than one for health, suggesting diseconomies. For total food, fuel and equipment, the sum of the coefficients is significantly different from one at the .06 level; these can therefore be regarded as marginal cases.

The objection to deflation by household size on the grounds of economies or diseconomies of scale does not thus appear to be serious in our sample. But the problem of spurious correlation, introduced by this procedure, still remains.

Comparison with Howe's Estimates

In 1963 Charles W. Howe prepared a set of income elasticities based on Nairobi survey data.¹ Some of these estimates differ sharply from ours, and it thus appears worthwhile to discuss the difference between Howe's study and ours.

Howe's estimates were obtained from the published grouped results of the Nairobi 1963 survey which we have used, combined with the Nairobi 1957/58 survey which covered lower income groups. The inclusion of the earlier lower income survey groups may explain part of the difference between Howe's figures and ours. Also, Howe used only 11 groups, compared with the 60 groups used in our study. The reliability of the estimates naturally increases with the number of groups used.

1. C.W. Howe, The Use of Sample Household Surveys in Economic Planning with Some Empirical Results for East Africa, Discussion Paper No. 12, Institute for Development Studies, Nairobi. Some of these results also appear in C. W. Howe, The Use of Sample Household Expenditure Surveys in Economic Planning in East Africa, Bulletin of the Oxford University Institute of Economics and Statistics, 1966.

Howe has estimated income rather than expenditure elasticities, using ordinary least squares. If our contention that income is measured with error is correct, then Howe's figures are questionable because one of the assumptions underlying the use of least squares is violated. Moreover, Howe deflated both expenditure and income by household size, introducing the possibility of spurious correlation, which can seriously distort the results. For a few items, the presence of economies or diseconomies of scale also makes this procedure questionable.

In Table 7 we present Howe's figures together with our estimates from the first set of regressions, the double log regressions with total expenditure and household size as explanatory variables. The items marked with an LL are those for which Howe used a double log function. For the other expenditure items, he used different functions, linear, semilog, log-inverse, etc. Howe does not indicate on what basis his choice between functional forms was made. But it is somewhat objectionable to make this choice simply on the basis of goodness of fit, and to use different functions for different expenditure items.

In the case of milk and eggs, and also meat and meals away, which Howe treated as single items, our figures are weighted averages of the estimated elasticities. Although the double log function is not additive, the average elasticities thus obtained are unlikely to differ markedly from the estimates one would obtain directly.

For the purposes of comparison our expenditure elasticities have been converted into income elasticities. If the elasticities are to be used for projecting demand, then an income elasticity of total expenditure of unity is a reasonable assumption, and our expenditure elasticities can also be regarded as income elasticities. For comparison on a cross-section basis, however, a total expenditure elasticity of less than one is a more appropriate assumption. We have presented three alternatives, assuming income elasticities of total expenditure of 1.0, 0.9 and 0.8. We feel it inadvisable to estimate the income elasticity of total expenditure from the original data, given that the income figures are unreliable.

Comparing our first set of figures, in column 2, with Howe's figures, we see that our estimates are higher in 11 cases, lower in the remaining 6. Comparing our column 3 figures with Howe's figures, our estimates are higher in 10 of the 17 cases. Our column 4 results are higher in 7 of the 17, when compared with Howe's estimates.

Howe appears to have underestimated the elasticities for cereals, sugar, meat/meals away, vegetables/fruit, total food and clothing, relative to our estimates. Similarly, his elasticities for tea/coffee, transportation, recreation, health, household operation and school fees are overestimated relative to ours. The estimates for the remaining items are close and depend on which of our 3 columns is used for the comparison. On balance, Howe appears to have overestimated the elasticities for foods and underestimated the elasticities for nonfood items.

In a few cases, the discrepancy between Howe's figures and ours is very large. In the case of transportation, this is due to the fact that Howe did not delete the extreme observation due to one household's purchase of a car in the survey month. This biases his estimate sharply upward. For school fees, Howe used the monthly figure, while we used one twelfth of the annual figure. Moreover, school fees is one of the few items for which there appear to be significant economies of scale. Deflating, as Howe does, by household size would thus bias this expenditure elasticity sharply upward. We find it difficult to believe that the school fees elasticity exceeds unity, too. Even among rural households with much lower incomes, in the Central Province, the elasticity of expenditure on school fees appears to be less than one.¹

Other major discrepancies exist for household operation and health, and for total food and clothing. We can not provide any obvious explanation for this, and conclude that the discrepancies are due to the different methods of estimation used. We feel that these large discrepancies cast doubt on the usefulness of the Howe results, and suggest that it is worthwhile to use the more difficult but theoretically more satisfactory methods of our study.

1. B. F. Massell, Determinants of Household Expenditure in Rural Kenya, Discussion Paper No. 49 Institute for Development Studies, Nairobi.

Demand Projections

We now present some very tentative demand projections, to illustrate the way in which our elasticities can be used:

We have already noted that our estimates cannot be assumed to apply to all income groups in Nairobi, nor to all urban groups in Kenya, but that until more data are available it is likely that they will be used for urban Kenya as a whole. There is a further problem which makes it necessary to treat our demand projections as tentative at this stage. Demand projections necessarily involve projected changes in income over time. It is hazardous to base them directly on income or expenditure elasticities estimated from cross-section data. It is common knowledge, for example, that although the income elasticity of saving estimated from cross-section data is invariably greater than one, the savings ratio tends to remain constant over time. Elasticities estimated from cross-section data tend to be long-run elasticities, whereas the elasticities appropriate for examining changes in income over short periods of time are short-run. And, cross-section estimates are affected by factors peculiar to the household, which remain constant over time. For these reasons, it is preferable to estimate income elasticities from re-interview data, which consist of repeated observations on the same sample of households over a period of time. Re-interview data are not available in Kenya, as yet, but it is as well to be aware of this additional reason for treating our results with caution at this stage.

Let us assume that household size will remain constant but that both the number of households and income per household will increase at constant percentage rates. Then the rate of increase in household demand, d_i , for an item i is given by

$$d_i = p + ye_i \quad (4)$$

where p = the rate of population increase,

y = the rate of increase in per household income, and

e_i = the income elasticity of demand for item i .

The income elasticity, e_i , is the product of the expenditure elasticity for item i and the income elasticity of total expenditure. If we assume that households save a constant proportion of their income as income rises, as has been observed to be the case in many countries over long periods of time, then the intertemporal income elasticity of total expenditure is unity, and income and expenditure elasticities are identical. Although it would be unrealistic to assume that the cross-section income elasticity of total expenditure is unity for a group of households at a point in time, it is decidedly more realistic to assume that the income elasticity of total expenditure is unity with respect to changes in income over time. We assume, then, that increases in total expenditure keep pace with increases in income, but that the allocation of expenditure among items changes as income rises, in the way estimated from the cross-section data.

For this demand projection exercise we use the first set of elasticities, those estimated from the double log function with total expenditure and household size as the explanatory variables. These expenditure elasticities are net of changes in household size, which we are assuming constant over time. The use of these estimates assumes either that other explanatory variables, such as education, retain the same relationship with total expenditure as is found in the sample, or that the other variables are not important in determining expenditure patterns.

Our tentative demand estimates are presented in Table 8. For the purposes of illustration, we have assumed a 4% rate of increase in per capita income, and two alternative rates of population growth: 3% in the first column, 6% in the second. With a 6% rate of urban population growth, the rates of increase in demand range from 13.8% for equipment to 5.5% for maize. If the rate of population growth is only 3%, the range is from 10.8% to 2.5%. Changes in the assumed rate of increase in population (or per capita income) do not alter the ordering of the items, but they do change the values, of course.

The smaller the assumed value of p relative to that of y , the more sensitive the results to differences in income elasticities. For example, when one assumes the rate of increase in population and in per capita income to be 3 and 4 percent, respectively, as in column 1 of the table, the rate of increase in demand for equipment is 4.3 times as great as that for maize. But when one assumes the rates of increase to be 6 and 4 percent, as in column 2 of the table, the rate of increase in demand for equipment is only 2.5 times that for maize.

These figures are based on the double log estimates. Using this function, projections are easily made, as the relative rate of increase in demand for each item is constant over time. In the ratio semilog estimates, however, elasticities decline as income rises, so that the relative rates of increase in demand decline over time. Thus calculations based on the semilog estimates are more complex.

Suggestions for Future Surveys

If the Kenya Government accepts the view that demand projections are a worthwhile exercise, and that such projections can usefully be based on budget study data, then it may be worth incorporating several changes in future household surveys. Probably the most rewarding change would be the introduction of re-interview sampling. It should prove feasible to select a group of households to be visited periodically over a number of years. As we have indicated above, the resulting data would be of far greater use as a basis for projecting demand patterns. By including price as well as income data, it would be possible to estimate both price and income elasticities.

A second suggestion, also of considerable importance, is to broaden the base of the survey. To make reliable national demand projections, it is important to have information representative of all areas of the country. Thus, one should have data covering the important income groups in Nairobi, smaller towns and townships and rural areas. Some data already exist for other areas in Kenya. The most ambitious undertaking has been the Central Province Survey, which includes 1080 rural and 360 urban households. However, for the purposes of demand projections, what is important is to have a single, broadly-based sample, permitting a uniform survey of households representative of the population as a whole.

The information collected could also be improved. The household is usually regarded as the basic unit with respect to expenditure and saving decisions. For this reason it is of considerable importance to obtain reliable information on total household income -- not only the income of the respondent. We feel that the irregular recording of income of other members of the household in the 1963 Nairobi Survey was a serious shortcoming.

Another point concerns the definition of expenditure. The distinction between purchases for cash and purchases for credit is an artificial one, which renders the data less useful. For the purposes of economic analysis it would be far better to include purchases on credit as well as those paid for in cash.

In general, it would be highly desirable to obtain more comprehensive information on saving, credit, and debt. Virtually nothing is known about consumer saving, its magnitude, and its disposition. Nor is there much information on consumer debt. It would be interesting, and useful for planning, to obtain information on the extent of debt, and to be able to analyse the relationship between consumer debt and such variables as purchases of durable goods, income, and ^{demographic} variables such as household size, age structure of the household, and educational level.

With regard to these demographic variables also, somewhat more comprehensive coverage could usefully be obtained. For example, although ages of various members of the household were obtained, no information was obtained on the age of the respondent. The respondent's educational level was ascertained, but no information was gathered on the education of other members of the household. These additional factors could provide valuable additional information on the determinants of expenditure patterns.

A final point concerns housing. This is admittedly a difficult item to handle. But, if possible, it would be useful to obtain information on the economic rent of housing occupied by each household. As we noted above, this should include any form of subsidy provided by the employer, the City Council or any other body. The full economic rent should then be included in the measure of household income as well.

R E F E R E N C E S

1. A. P. Barten, H. Theill, and C. T. Leenders, "Farmers Budgets in a Depression Period," Econometrica, 1962, pp. 546-564.
2. Robert Ferber, "Research on Household Behavior," American Economic Review, March 1962.
3. Milton Friedman, A Theory of the Consumption Function, Princeton University Press, 1957.
4. Irwin Friend and Robert Jones (editors), Study of Consumer Expenditures, Incomes, and Savings. Proceedings of the Conference on Consumption and Saving. University of Pennsylvania, Wharton School of Finance and Commerce, 1960.
5. Donald J. Harris, "Econometric Analysis of Household Consumption in Jamaica," Social and Economic Studies, pp. 471-487.
6. H. S. Houthakker, "An International Comparison of Household Expenditure Patterns, Commemorating the Centenary of Engel's Law," Econometrica, Vol. 25, 1957, pp. 532-551.
7. H. S. Houthakker, "New Evidence on Demand Elasticities," Econometrica, April 1965, pp. 277-288.
8. Charles W. Howe, "The Use of Sample Household Expenditure Surveys in Economic Planning in East Africa," Bulletin of the Oxford University Institute of Economics and Statistics, Vol. 28, No. 3, 1966, pp. 199-209. (An earlier version, but with additional estimates, published as Discussion Paper No. 12, Institute for Development Studies, Nairobi, 1963, mimeo.)
9. Kenya Government, Directorate of Planning, Ministry of Finance and Economic Planning, The Pattern of Income, Expenditure, and Consumption of African Middle Income Workers in Nairobi, July 1963. Nairobi, July 1964.
10. Edwin Kuh, "The Validity of Cross-sectionally Estimated Behavior Equations in Time Series Applications," Econometrica, April 1959, pp. 197-214.
11. C. E. V. Leser, "Forms of Engle Functions," Econometrica, 1963, pp. 694-703.
12. N. Liviatan, "Errors in Variables and Engel Curve Analysis," Econometrica, 1961, pp. 336-362.

13. A. Madansky, "The Fitting of Straight Lines When Both Variables are Subject to Error," Journal of the American Statistical Association, Vol. 54, 1959.
14. A. Madansky, Spurious Correlation Due to Deflating Variables, The RAND Corporation, Memorandum RM-3863-PR, December 1963.
15. H. W. Ord, Ghana - Projected Level of Demand, Supply, and Imports of Agricultural Products in 1965, 1970, and 1975. Economic Research Service, Foreign Agricultural Service, U. S. Department of Agriculture, 1964.
16. S. J. Prais and H. S. Houthakker, The Analysis of Family Budgets, Cambridge University Press, 1955.
17. Lewis Schipper, Consumer Discretionary Behavior, North Holland Publishing Co., Amsterdam, 1964.
18. R. Stone, The Measurement of Consumers' Expenditure in the United Kingdom, 1920-1938, Vol. 1, Cambridge, 1954.
19. R. Summers, "A Note on Least Squares Bias in Household Expenditure Analysis," Econometrica, Vol. 27, 1959.